MathWorks Math Modeling Challenge 2019

High Technology High School–

Team # 12038 Lincroft, New Jersey

Coach: Raymond Eng

Students: Eric Chai, Gustav Hansen, Emily Jiang, Kyle Lui, Jason Yan

MathWorks Math Modeling Challenge Champions

$20,000 Team Prize

MathWorks Math Modeling Challenge

Substance Use and Abuse

Executive Summary

In recent years, substance abuse has intensified to an alarming degree in the United States. In particular, the rise of vaping, a new form of nicotine consumption, is dangerously exposing drug abuse to a new generation. With the need to understand how substance use spreads and impacts individuals differently, our team seeks to provide a report with mathematically founded insights on this prevalent issue.

We first strove to predict the spread of nicotine use due to both vaping and cigarettes over the next decade. By comparing the spread of nicotine use to an infectious disease, we modified the SIRS epidemiology model to create our adapted SIRI model in which individuals are divided into four compartments: infected (drug users), recovered (users who quit drugs), susceptible (potential drug users), and nonsusceptible (those who will never use drugs). People progress from susceptible to infected to recovered, but may relapse into their old habits, causing them to re-enter the infected population. Birth and death rates of our designated population were modeled with linear equations. We solved a system of differential equations to determine e-cigarette and cigarette use in 2029: 26.63% of the American population will vape and 6.45% will smoke cigarettes. These results align with the expectation that vaping will increase in popularity while cigarette smoking will decline.

Substance abuse is associated with numerous social factors and personal attributes. We incorporated those determinants to create a second mathematical model that computes the probability that an individual will use nicotine, marijuana, alcohol, and unprescribed opioids. A binary multivariate logistic model was used to assess the effects of age, gender, ethnicity, income, parental status, friendship, opinion about school, overall health, weapon possession, and bullying on substance use. To demonstrate our model, we coded and executed a Monte Carlo simulation that created 300 high school seniors with varying attributes. We found that 46.3% of the students would use nicotine, 17.3% would use marijuana, 66.0% would use alcohol, and 0.0% would use opiates.

Substance use has far-reaching implications in personal and societal spheres. It is crucial to rank substances based on their overall impact in order to assess necessary government action regarding drug abuse. To address this issue, we developed a robust metric to rank the effects of nicotine, marijuana, alcohol, and opioid abuse. Our model and ranking considers physical harm, dependence, social harm, and economic impact of the drugs. The former three factors were measured on a scale of 0 to 3 based on psychiatrist surveys. Then economic impact was defined as GDP loss from the decrease in life expectancy caused by drug abuse. After applying risk factors obtained from the amount of people that use each drug, the four substances were ranked. From highest to lowest individual impact, the ranking was opioids, alcohol, cigarettes, and marijuana. From highest to lowest total societal impact, the ranking was alcohol, cigarette, marijuana, and opioids.

The repercussions of substance abuse are reverberating and remain with an individual for life. However, drugs not only severely affect the user but also cause extensive societal harm. Increased understanding of the projected spread and impact of substance abuse, as well as the underlying factors that lead to poor judgment, are needed to optimize measures to restrict consumption. Ultimately, we believe that our models provide novel insight into the nationwide issue of substance use and abuse.

1 Introduction

This section delineates the components of the modeling problem and their objectives. Global assumptions applying to the entire modeling process are also listed.

1.1 Restatement of the Problem

The problem we are tasked with addressing is as follows:

1. Build a mathematical model that predicts the spread of nicotine use due to vaping over the next 10 years. Analyze how this growth compares to that of cigarettes.

2. Create a model that simulates the likelihood that a given individual will use a given substance, accounting for social influence, characteristic traits, and properties of the drug itself. Demonstrate the model by predicting how many students among a class of 300 high school seniors with varying characteristics will use nicotine, marijuana, alcohol, and unprescribed opioids.

3. Develop a metric for the impact of substance use, considering both financial and nonfinancial factors. Use the metric to rank the substances listed in Part II.

1.2 Global Assumptions

1. *The current drug scene remains constant.* We assume that there will be no radical changes in the recreational drug industry, such as new drugs or drug products. This assumption is imperative because attempting to account for unpredictable and volatile factors would make model development virtually impossible.

2. *All vapes count as e-cigarettes.* Some people distinguish between e-cigarettes and vaping. For the purposes of this model, e-cigarettes and vapes will be considered synonymous.

3. *People respond honestly to surveys.* Our model is dependent on survey results to calculate weight constants. Because we have no way of determining the accuracy of the survey responses, we will assume that they are accurate and without bias for simplicity.

2 Part 1: Darth Vapor

First commercialized in 2003, electronic cigarettes have become an increasingly popular product among youth [1]. Although they are advertised as safer alternatives to traditional cigarettes, e-cigarettes contain high doses of nicotine and have introduced a new generation to tobacco products. This section outlines a mathematical model for predicting the change in nicotine use in the United States due to vaping compared to the change due to cigarettes.

2.1 Assumptions

1. *Nicotine use can be modeled as an infectious disease.* Like an epidemic, nicotine use is prevalent and contagious, reflected in the surge in popularity of smoking due to peer pressure, advertisements, and social media. Additionally, the U.S. Surgeon General declared youth vaping a nationwide epidemic in 2018 [2].

2. *Individuals can smoke from age 11 until death.* Peak years for first trying nicotine products is 6th or 7th grade [3].

3. *Rate of entry into pre-adolescence in the U.S. is 0.00103.* [4] Our model defines "birth" as reaching an age at which substance use becomes possible—around 11 years. Thus, we assumed the current birth rate to be constant for the past 11 years, assuming no children die before they turn 11. The current birth rate is 1.03 people/month/person.

4. *Death rate in the U.S. is constant and equal to 0.0007 people per month per person.*[4] Our model assumes that individuals have the capacity to use drugs until their death.

5. *Individuals can only start smoking due to influence from other smokers.* To model substance use as an infectious disease, we must assume that susceptible individuals can become infected only from contact with the already infected. This assumption is valid because peer influence and social media presence are the driving factors behind the popularity of smoking [5].

6. *Individuals are either not susceptible to, susceptible to, infected by, or recovered from substance abuse.* As in the SIR epidemiology model, we assume that people are either unwilling to smoke (not susceptible), open to smoking (susceptible), regular smokers (infected), or past smokers who have quit (recovered).

7. *The infection rate is constant over time.* Because we are assuming that the drug industry does not drastically change, it is reasonable to assume that the infection rate will also not drastically change.

8. *The percentage of susceptible people will stay constant over time.* Because we are assuming that the drug industry does not drastically change, it is reasonable to assume that the number of people susceptible to it will also not drastically change.

9. *Nobody starts as recovered.* At the start of the model, we do not consider any individuals to be former smokers who have quit.

10. *The recovery and relapse constant for cigarette and e-cigarette users are the same.* The two contain similar amounts of nicotine, which acts as the addictive agent. Thus, the recovery and relapse constants are assumed to be the same.

2.2 Model Development

The surge in popularity of conventional cigarettes in the mid-20th century, as well as the current boom of vaping among American youth, is comparable to the spread of an infectious disease during an epidemic. As stated in assumption 1, we model nicotine use as a disease because it rapidly spreads as a result of interpersonal communications (in-person peer pressure to try a drug as well as social media prevalence); additionally, substance use is a condition from which individuals can recover (by quitting smoking).

Our model is a derivation of the SIRS epidemiological model, a technique used to map the spread of infectious diseases such as influenza. We also consider birth and death rate, since population naturally changes over time. The model separates individuals in a population into four categories: NS for Not Susceptible, S for Susceptible, I for Infected, and R for Recovered. At the start of the model, individuals are either in NS, S, or I, since nobody starts off as recovered. While those in NS remain there permanently, individuals in S can move to I, who can then move to R.

The additional S in SIRS represents the possibility of returning to the Susceptible compartment—in this case, a regular user quitting but relapsing. However, we modified the classic SIRS model by recognizing that a relapsing individual would re-enter the Infected category rather than Susceptible, since they will once again become smokers rather than people merely open to smoking. Thus, we renamed the traditional epidemiology model as SIRI to represent this adjustment. Figure 2.2.1 diagrams the aforementioned movement of individuals between categories, while Table 2.2.1 defines and details values for variables and constants used in the SIRI model for both e-cigarette and cigarette smoking.

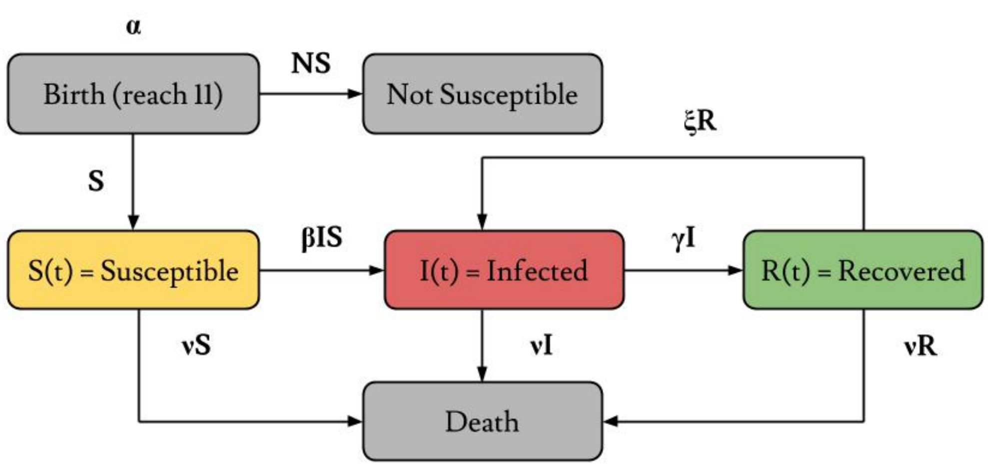

Figure 2.2.1: Diagram of the SIRI Model for Spread of Nicotine Use

2.2.1 Parameters in SIRI Model

Proportion of infected people (I_0). The total number of people that currently vape is approximately 10.8 million [6]. Dividing by the total population of America, 325.7 million

[7], results in an I_0 value of 0.0332 for e-cigarettes. The total number of people that currently smoke cigarettes is approximately 34.3 million [8], resulting in an I_0 value of 0.1053.

Proportion of recovered people (R_0). As per assumption 9, without loss of generality, R_0 was assumed to be 0 at time = 0.

Proportion of susceptible people (S_0). Because I, R, and S are proportions of the total population, their sums must add to 1. Thus, $S_0 = 1 - R - I$, resulting in 0.9667 for e-cigarettes and 0.8947 for cigarettes.

Susceptibility (S). A 2016 Surgeon General report stated that 32% of people are considered susceptible to e-cigarette use [5], while a 2012 report stated that 20% of people are susceptible to cigarettes, which correspond to the S values [9].

Infection constant (β). This was determined based on responses to the survey question "If one of your best friends offered you a cigarette, would you smoke it?" For e-cigarettes, the chance of infection was taken from a 2016 U.S. Surgeon General report that indicated that 18% of young adults responded "yes" to the question [5]. For cigarettes, we obtained β by adding the percentages of the responses "Definitely Yes" and "Probably Yes," from the 2014 National Survey on Drug Use and Health, to get 0.3%, which represented the infection constant [10].

Recovery constant (γ). In a given year, around 40% of smokers attempt to quit [11]. Therefore, in a month, $1.40^{1/12} = 1.0284$ recover, so the recovery rate is 0.0284.

Relapse constant (ξ). In a given year, approximately 6% of attempts to quit smoking succeed and 94% of attempts failed and the person relapsed [12]. Therefore, in a month, $1.94^{1/12} = 1.0568$ fail, so the relapse constant is 0.0568.

Infection rate (y_{inf}). In accordance with assumption 4, we assume that people will only start smoking if they are influenced by a current smoker. In other words, a susceptible person can only become infected if they come into contact with an infected person, which occurs at a rate proportional to $I \cdot S$. The infection constant β represents the likelihood that a susceptible person becomes infected when influenced by a smoker. Thus, infection rate is as follows:

$$y_{inf} = \beta \cdot I \cdot S \tag{1}$$

Recovery rate (y_{rec}). Unlike infection rate, the recovery rate is dependent only on the average probability of an individual quitting. The recovery constant γ multiplied by the proportion of people that currently are infected gives the recovery rate:

$$y_{rec} = \gamma \cdot I \tag{2}$$

Relapse rate (y_{rel}). The relapse rate is dependant only on the average probability of an individual relapsing. The relapse constant is much higher than the infection rate, which is logical because an individual who was previously a regular smoker will be more likely to succumb to the addictive cycle again [12]. Designating ξ as the relapse constant, relapse rate is given by

$$y_{rel} = \xi \cdot R \tag{3}$$

Birth rate (α). The birth rate, as defined by assumption 3, is 1.03 people/month/person.

Death rate (μ). From assumption 4, the death rate is assumed to be constant and equal to 0.0007 people per month per person. Therefore, the number of people dead for each category will be the death rate multiplied by the proportion of the people in each category.

$$\mu_S = v \cdot S \tag{4}$$

$$\mu_I = v \cdot I \tag{5}$$

$$\mu_R = v \cdot R \tag{6}$$

Table 2.2.1 Variables and Constants of SIRI Model for E-Cigarettes and Cigarettes

Variable	Definition	E-Cigarette Values	Cigarette Values
I	Proportion of infected people	$I_0 = 0.0332$	$I_0 = 0.1053$
R	Proportion of recovered people	$R_0 = 0$	$R_0 = 0$
S	Proportion of susceptible people	$S_0 = 0.9667$	$S_0 = 0.8947$
N	Proportion of total individuals in SIR cycle	$N_0 = 0.32$	$N_0 = 0.20$
α	Birth rate	0.00103	0.00103
β	Infection constant	0.18	0.003
γ	Recovery constant	0.0284	0.0284
ξ	Relapse constant	0.0568	0.0568
μ	Death rate	0.0007	0.0007

2.2.2 Differential Equations for SIRI Model

The change in each of the dependent variables S, I, and R is equal to the sum of the input of the respective category minus the sum of its output, as diagrammed by the arrows entering and leaving each box in Figure 2.2.1. Thus, our SIRI model is summarized by the set of ordinary differential equations below:

$$\frac{dS}{dt} = \alpha - \beta \cdot I \cdot S - \mu \cdot S \tag{7}$$

$$\frac{dI}{dt} = B \cdot I \cdot S - \gamma \cdot I + \xi \cdot R - \mu \cdot I \tag{8}$$

$$\frac{dR}{dt} = \gamma \cdot I - \xi \cdot R - \mu \cdot R \tag{9}$$

2.3 Results

With the SIRI model established, we utilized it to predict the change in nicotine use due to e-cigarettes and cigarettes in the next decade. We coded and executed a Python program to solve the system of differential equations, with appropriate constants for each product, and graph the proportion of compartments over time. Figures 2.3.1 and 2.3.2 graph the proportion of the total population falling under each of the SIR categories for both tobacco products, respectively, over a 10-year time period. Table 2.3.1 enumerates the proportion of the population that is susceptible, infected, and recovered for vaping and cigarettes in 2029.

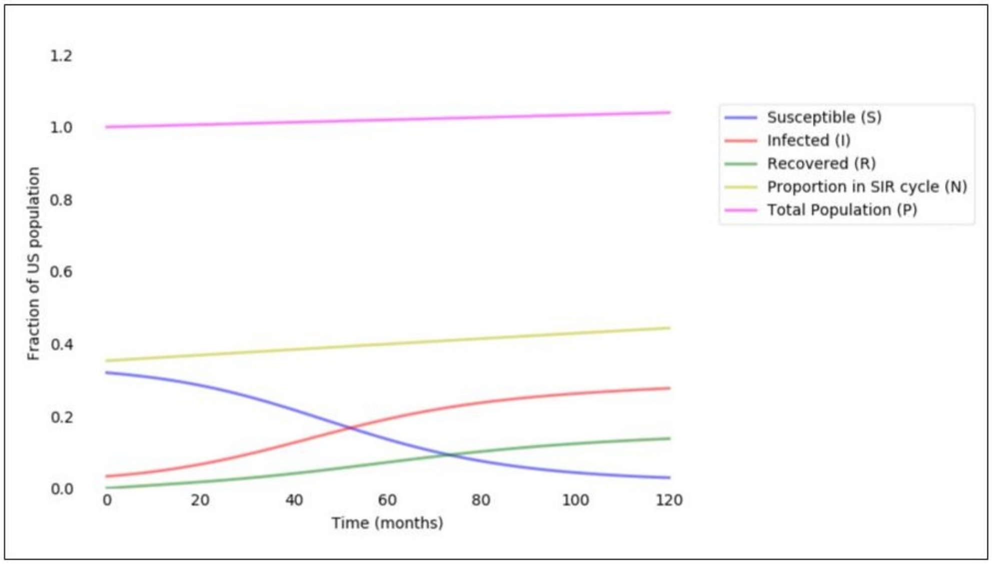

Figure 2.3.1: Graph of SIRI Compartments for E-Cigarettes over Ten Years

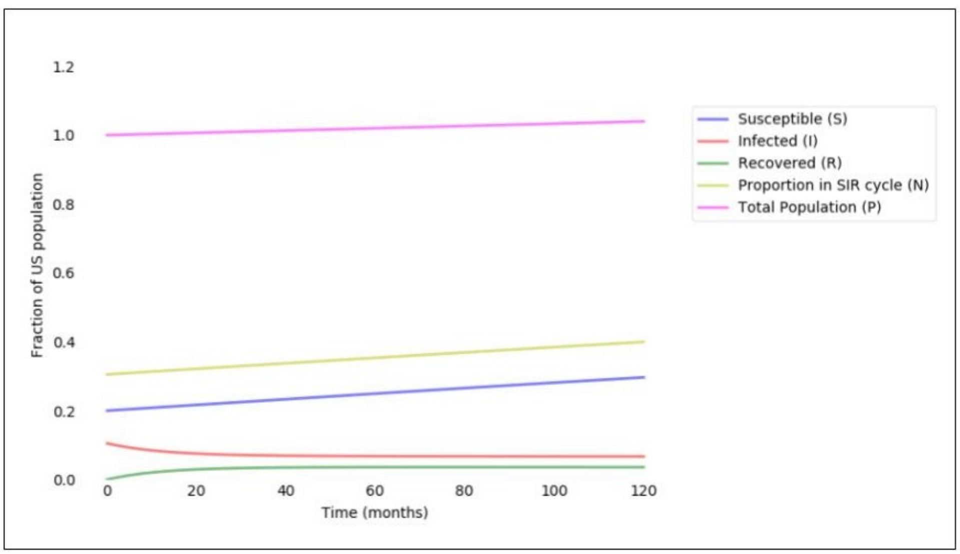

Figure 2.3.2: Graph of SIRI Compartments for Cigarettes over Ten Years

Table 2.3.1 SIR Distribution of 2029 Population for E-Cigarettes and Cigarettes

	Susceptible	Infected	Recovered
E-Cigarettes	2.82%	26.63%	13.21%
Cigarettes	28.53%	6.45%	3.45%

Our model concludes that in 2029, 26.63% of the population will use e-cigarettes, while 6.45% will use cigarettes. This disparity is consistent with previously researched trends, which suggest that as e-cigarettes gain popularity amongst teens, regular cigarettes decrease in popularity [13].

2.4 Sensitivity Analysis

Table 2.4.1 shows the sensitivity analysis for our SIRI model based on an independent increase and decrease of 10% of the infection constant β, recovery constant γ, and relapse constant ξ.

Table 2.4.1 Sensitivity Analysis for Part I

Constant	% Change in Constant	% Change in Vaping (I)	% Change in Cigarette Use (I)
β	10%	1.014%	0.6202%
β	-10%	-1.615%	-0.6202%
γ	10%	-3.492%	-3.566%
γ	-10%	4.018%	3.721%
ξ	10%	3.098%	3.367%
ξ	-10%	-3.496%	-3.905%

Positive changes in the infection or relapse constants resulted in positive changes in the percentage of infected people for both vaping and cigarette use. This is consistent with our predictions because the rate of infection for susceptible and recovered people is increasing. In contrast, a positive change in recovery constant resulted in a decrease in percent infected because the rate at which people are leaving the infected population is increasing.

2.5 Strengths and Weaknesses

Our model is resilient to small changes and outputs sensible results. As demonstrated in the sensitivity analysis, a 10% change in each of the infection, recovery, and relapse constants accounts for less than 5% change in final vaping and cigarette use after a decade. Changes in the model's output due to shifts are consistent with expected trends as well. SIRS is also an established mathematical modeling technique that we adapted to fit our own aims, lending credence to the validity of our model. Additionally, our model is comprehensive, accounting for many contributing factors such as population change, nonsusceptible individuals, and the possibility of relapse for smokers who have attempted to quit.

The model's weaknesses lie in its inability to account for the introduction of new forms of drugs or rapid changes in popularity of existing forms, as stated in global assumption 1. Specifically, a surge in use of a particular drug would likely impact vaping and cigarette use in unforeseen ways that our model will not accurately predict. Furthermore, our model does not consider the association between vaping and cigarette use, and how the growth or decline of one product would influence the other. This is unrealistic because the popularity of e-cigarettes among youth has led many to smoke traditional cigarettes and prompted cigarette smokers to transition to vaping [13]; however, the opposite effects of these two phenomena can reasonably counterbalance each other.

5 Conclusion

5.1 Further Studies

Our first model does not currently account for the introduction of new drugs in the industry, which would greatly impact the change in usage for pre-existing substances. Taking these market changes into account would greatly strengthen our model. The second model used survey data from 2005–2006. The resulting model fits well for this time period, but requires more recent data to reflect recent trends. Applying the same modeling approach for 2019 would create a more accurate model that is applicable to today. Finally, the third model is heavily based on the personal opinions of psychiatrists. Recreating the model to account for each factor with independent methods would greatly complicate the model, but make it more flexible for incorporating newer drugs into our ranking.

5.2 Summary

The first model focuses on comparing the percent of e-cigarette users versus cigarette users in the next ten years. The SIRS epidemic model was used as the basis for ours. People were split into four main categories: infected (those that used drugs), recovered (those that quit using drugs), susceptible (those that may use drugs in the future), and non-susceptible (those that will never use drugs). Birth rate and death rate were both modeled with linear equations. Simultaneous differential equations were solved to determine the number of "infected" people in 2029. According to our model, 26.63% of the American population will vape in 2029 and 6.45% will smoke cigarettes. The results correspond with observed increasing popularity of e-cigarettes and decreasing popularity of regular cigarettes.

The second model determines the probability of a student using nicotine, marijuana, alcohol, and opioids and applies itself to a randomly generated sample of 300 high school seniors. A binary multivariate logistic regression was used to create the model based on an HBSC survey. A machine learning algorithm using an L2 regression was used to calculate the weights and bias in our logistic model. Using a Monte Carlo simulation, 300 random seniors were created based on response frequencies to each of questions necessary for our model. Running this sample of high school seniors through our model, we found 46.33% would use nicotine, 17.33% would use marijuana, 66.00% would use alcohol, and 0.00% would use opiates.

The third and final model focuses on ranking nicotine, marijuana, alcohol, and opioids based on their financial and nonfinancial effects. Factors were analyzed in four main categories: physical harm, dependence, social harm, and economic impact. These factors were further split into 2–3 subcategories each that were each assigned scores on a scale from 0.0 to 3.0 based on expert surveys. To calculate the impact of drugs on GDP, the average annual GDP per person was multiplied by the average decrease in life as a result of using drugs. The impact of drugs on GDP was then rescaled from 0.0 to 3.0 to make them comparable to the other factors. Each of the four main categories was averaged for a total harm score for each of the four drugs. The total harm score was multiplied by a risk factor based on the number of people that used each drug to obtain a final score for each drug that could be used for ranking purposes. This model showed that opioids had the greatest substance harm per person, but since relatively few people use opiods, it had a lower total detriment score. Marijuana had the lowest substance harm per person and the second lowest total impact. Alcohol had the highest total impact, while cigarettes had the second highest because of the great number of people using these substances.

6 References

[1] - Historical Timeline of Electronic Cigarettes. (2018, October 18). Retrieved March 3, 2019, from http://www.casaa.org/historical-timeline-of-electronic-cigarettes/

[2] - Stein, R. (2018, December 18). Surgeon General Warns Youth Vaping Is Now An 'Epidemic'. Retrieved March 3, 2019, from https://www.npr.org/sections/health-shots/2018/12/18/677755266/surgeon-general-warns-youth-vaping-is-now-an-epidemic

[3] - Bach, L. (2018). The Path to Addiction Starts Early. Retrieved March 3, 2019, from https://www.tobaccofreekids.org/assets/factsheets/0127.pdf

[4] - Birth rate, crude (per 1,000 people). (2019). Retrieved March 3, 2019, from https://data.worldbank.org/indicator/SP.DYN.CBRT.IN?locations=US&view=chart

[5] - United States, U.S. Department of Health and Human Services, Office on Smoking and Health. (2016). Surgeon General. Retrieved March 3, 2019, from https://e-cigarettes.surgeongeneral.gov/documents/2016_SGR_Full_Report_non-508.pdf

[6] - Mirbolouk, M., Charkhchi, P., Kianoush, S., Uddin, S. I., Orimoloye, O. A., Jaber, R., . . . Blaha, M. J. (2018). Prevalence and Distribution of E-Cigarette Use Among U.S. Adults: Behavioral Risk Factor Surveillance System, 2016. Annals of Internal Medicine, 169(7), 429. doi:10.7326/m17-3440

[7] - U.S. Population (LIVE). (n.d.). Retrieved March 3, 2019, from http://www.worldometers.info/world-population/us-population/

[8] - Current Cigarette Smoking Among Adults in the United States — CDC. (n.d.). Retrieved March 3, 2019, from https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm

[9] - United States., Public Health Service., Office of the Surgeon General. (2012). A report of the Surgeon General: Preventing tobacco use among youth and young adults. Washington, D.C.?: Centers for Disease Control and Prevention (U.S.), Office on Smoking and Health.

[10] - Substance Abuse and Mental Health Services Administration. (2014). 2014 National Survey On Drug Use And Health [Public Use File Codebook]. Retrieved from https://samhda.s3-us-gov-west-1.amazonaws.com/s3fs-public/field-uploads-protected/studies/NSDUH-2014/NSDUH-2014-datasets/NSDUH-2014-DS0001/NSDUH-2014-DS0001-info/NSDUH-2014-DS0001-info-codebook.pdf

[11] - Borland, R., Partos, T. R., Yong, H., Cummings, K. M., & Hyland, A. (2012). How much unsuccessful quitting activity is going on among adult smokers? Data from the International Tobacco Control Four Country cohort survey. Addiction, 107(3), 673-682. doi:10.1111/j.1360-0443.2011.03685.x

[12] - Malarcher, A., & Dube, S. (2011). Quitting Smoking Among Adults — United States, 2001–2010 (United States, Centers for Disease Control and Prevention).

[13] - National Institute on Drug Abuse. (2016, February 11). Teens and E-cigarettes. Re-

trieved March 3, 2019, from https://www.drugabuse.gov/related-topics/trends-statistics/infographics/teens-e-cigarettes

[14] - Martinez, E., Kaplan, C. P., Guil, V., Gregorich, S. E., Mejia, R., & Pérez-Stable, E. J. (2006). Smoking behavior and demographic risk factors in Argentina: A population-based survey. Prevention and Control, 2(4), 187-197. doi:10.1016/j.precon.2007.07.002

[15] - Year / Grade Placement. (n.d.). Retrieved March 3, 2019, from https://www.acs-schoo com/acs-hillingdon-class-ages

[16] - United States Department of Health and Human Services, Substance Abuse & Mental Health Data Archive. (2005-2006). Health Behavior in School-Aged Children. Retrieved from https://www.icpsr.umich.edu/icpsrweb/

[17] - Monitoring the Future National Survey Results on Drug Use, 1975-2005: Secondary School Students 2005. (2005). doi:10.3998/2027.42/142406

[18] - Golub, A., Johnson, B. D., & Dunlap, E. (2006). The Growth in Marijuana Use Among American Youths During the 1990s and the Extent of Blunt Smoking. Journal of Ethnicity in Substance Abuse, 4(3-4), 1-21. doi:10.1300/j233v04n03_01

[19] - Felter, C. (2019, January 17). The U.S. Opioid Epidemic. Retrieved March 3, 2019, from https://www.cfr.org/backgrounder/us-opioid-epidemic

[20] - ACMT FAQ Nicotine. (n.d.). Retrieved March 3, 2019, fromhttps://www.acmt.net/_Library/Public_Affairs/ACMT_FAQ_Nicotine_.pdf

[21] - National Institute on Drug Abuse. (2014). Principles of Adolescent Substance Use Disorder Treatment: A Research-Based Guide. Retrieved March 3, 2019, from https://www.drugabuse.gov/publications/principles-adolescent-substance-use-disorder-treatment- research-based-guide/introduction

[22] - Nutt, D., King, L. A., Saulsbury, W., & Blakemore, C. (2007). Development of a rational scale to assess the harm of drugs of potential misuse. The Lancet, 369(9566), 1047-1053. doi:10.1016/s0140-6736(07)60464-4

[23] - COUNTRY COMPARISON :: GDP - PER CAPITA (PPP). (n.d.). Retrieved March 3, 2019, from https://www.cia.gov/library/publications/the-world-factbook/rankorder/2004rank.html

[24] - Substance Abuse and Mental Health Services Administration. (2018). Results From the 2017 National Survey on Drug Use and Health: Detailed Tables. Retrieved from https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHDetailedTabs2017/NSDUHDetailedTabs2017.pdf